

# Nearly Optimal NP-Hardness of Vertex Cover on $k$ -Uniform $k$ -Partite Hypergraphs

Sushant Sachdeva \*

Rishi Saket †

January 22, 2013

## Abstract

We study the problem of computing the minimum vertex cover on  $k$ -uniform  $k$ -partite hypergraphs when the  $k$ -partition is given. On bipartite graphs ( $k = 2$ ), the minimum vertex cover can be computed in polynomial time. For general  $k$ , the problem was studied by Lovász [23], who gave a  $\frac{k}{2}$ -approximation based on the standard LP relaxation. Subsequent work by Aharoni, Holzman and Krivelevich [1] showed a tight integrality gap of  $(\frac{k}{2} - o(1))$  for the LP relaxation. While this problem was known to be NP-hard for  $k \geq 3$ , the first non-trivial NP-hardness of approximation factor of  $\frac{k}{4} - \varepsilon$  was shown in a recent work by Guruswami and Saket [13]. They also showed that assuming Khot's Unique Games Conjecture yields a  $\frac{k}{2} - \varepsilon$  inapproximability for this problem, implying the optimality of Lovász's result.

In this work, we show that this problem is NP-hard to approximate within  $\frac{k}{2} - 1 + \frac{1}{2k} - \varepsilon$ . This hardness factor is off from the optimal by an additive constant of at most 1 for  $k \geq 4$ . Our reduction relies on the *Multi-Layered PCP* of [8] and uses a gadget – based on biased Long Codes – adapted from the LP integrality gap of [1]. The nature of our reduction requires the analysis of several Long Codes with different biases, for which we prove structural properties of the so called *cross-intersecting* collections of set families – variants of which have been studied in extremal set theory.

## 1 Introduction

A  $k$ -uniform hypergraph  $G = (V, E)$  consists of a set of vertices  $V$  and a collection of hyperedges  $E$  such that each hyperedge contains exactly  $k$  vertices. A vertex cover for  $G$  is a subset of vertices  $\mathcal{V} \subseteq V$  such that every hyperedge  $e$  contains at least one vertex from  $\mathcal{V}$  i.e.  $e \cap \mathcal{V} \neq \emptyset$ . Equivalently, a vertex cover is a hitting set for the collection of hyperedges  $E$ . The complement of a vertex cover is called an *Independent Set*, which is a subset of vertices  $\mathcal{I}$  such that no hyperedge  $e \in E$  is contained inside  $\mathcal{I}$  i.e.  $e \not\subseteq \mathcal{I}$ .

The  $k$ -HYPVC problem is to compute the minimum vertex cover in a  $k$ -uniform hypergraph  $G$ . It is an extremely well studied combinatorial optimization problem, especially on graphs ( $k = 2$ ), and is known to be NP-hard. Indeed, the minimum vertex cover problem on graphs was one of Karp's original 21 NP-complete problems [19]. On the other hand, the simple greedy algorithm that picks a maximal collection of disjoint hyperedges and includes all vertices in the edges in the vertex cover gives a  $k$ -approximation, which is also obtained by the standard LP relaxation of the problem. The best algorithms known today achieve only a marginally better approximation factor of  $(1 - o(1))k$  [18, 15].

On the intractability side, there have been several results. For the case  $k = 2$ , Dinur and Safra [9] obtained an NP-hardness of approximation factor of 1.36, improving on a  $\frac{7}{6} - \varepsilon$  hardness by Håstad [14]. For general  $k$

\*Department of Computer Science, Princeton University. sachdeva@cs.princeton.edu

†Department of Computer Science, Princeton University. rsaket@cs.princeton.edu

a sequence of successive works yielded improved NP-hardness factors:  $\Omega(k^{1/19})$  by Trevisan [27];  $\Omega(k^{1-\varepsilon})$  by Holmerin [16];  $k - 3 - \varepsilon$  by Dinur, Guruswami and Khot [7]; and the currently best  $k - 1 - \epsilon$  due to Dinur, Guruswami, Khot and Regev [8]. In [8], the authors build upon [7] and the work of Dinur and Safra [9]. Moreover, assuming Khot's Unique Games Conjecture (UGC) [20], Khot and Regev [21] showed an essentially optimal  $k - \varepsilon$  inapproximability. This result was further strengthened in different directions by Austrin, Khot and Safra [5] and by Bansal and Khot [6].

### *Vertex Cover on $k$ -uniform $k$ -partite Hypergraphs*

In this paper we study the minimum vertex problem on  $k$ -partite  $k$ -uniform hypergraphs, when the underlying partition is given. We denote this problem as  $k$ -HYPVC-PARTITE. This is an interesting problem in itself and its variants have been studied for applications related to databases such as distributed data mining [10], schema mapping discovery [11] and optimization of finite automata [17]. On bipartite graphs ( $k = 2$ ), by Kőnig's Theorem computing the minimum vertex cover is equivalent to computing the maximum matching which can be done efficiently. For general  $k$ , the problem was studied by Lovász who, in his doctoral thesis [23], proved the following upper bound.

**Theorem 1.1 (Lovász [23])** *For every  $k$ -partite  $k$ -uniform hypergraph  $G$ :  $\text{vc}(G)/\text{LP}(G) \leq k/2$ , where  $\text{vc}(G)$  denotes the size of the minimum vertex cover and  $\text{LP}(G)$  denotes the value of the standard LP relaxation. This yields an efficient  $k/2$  approximation for  $k$ -HYPVC-PARTITE.*

The above upper bound was shown to be tight by Aharoni, Holzman and Krivelevich [1] who proved the following theorem.

**Theorem 1.2 (Aharoni et al.[1])** *For every  $k \geq 3$ , there exists a family of  $k$ -partite  $k$ -uniform hypergraphs  $G$  such that  $\text{vc}(G)/\text{LP}(G) \geq k/2 - o(1)$ . Thus, the integrality gap of the standard LP relaxation is  $k/2 - o(1)$ .*

A proof of the above theorem describing the integrality gap construction is included in Section A. The problem was shown to be APX-hard in [17] and [11] for  $k = 3$  which can be extended easily to  $k \geq 3$ . A recent work of Guruswami and Saket [13] showed the following non-trivial hardness of approximation factor for general  $k$ .

**Theorem 1.3 (Guruswami and Saket [13])** *For any  $\epsilon > 0$  and  $k \geq 5$ ,  $k$ -HYPVC-PARTITE is NP-hard to approximate within a factor of  $\frac{k}{4} - \epsilon$ . Assuming the UGC yields an optimal hardness factor of  $\frac{k}{2} - \epsilon$  for  $k \geq 3$ .*

**Our Contribution.** We show a nearly optimal NP-hardness result for approximating  $k$ -HYPVC-PARTITE.

**Theorem 1.4** *For any  $\epsilon > 0$  and integer  $k \geq 4$ , it is NP-hard to approximate the minimum vertex cover on  $k$ -partite  $k$ -uniform hypergraphs within to a factor of  $\frac{k}{2} - 1 + \frac{1}{2k} - \epsilon$ .*

Our result significantly improves on the NP-hardness factor obtained in [13] and is off by at most an additive constant of 1 from the optimal for any  $k \geq 4$ . The next few paragraphs give an overview of the techniques used in this work.

**Techniques.** It is helpful to first briefly review the hardness reduction of [8] for  $k$ -HYPVC which begins with the construction of a new *Multi-Layered PCP*. This is a two variable CSP consisting of several *layers* of variables, and constraints between the variables of each pair of layers. The work of [8] shows that it is NP-hard to find a labeling to the variables which satisfies a small fraction of the constraints between *any* two layers, even if there is a labeling that satisfies all the constraints of the instance. The reduction to a

$k$ -uniform hypergraph (as an instance of  $k$ -HYPVC) involves replacing each variable of the PCP with a biased Long Code, defined in [9], where the bias depends on  $k$ .

The starting point for our hardness reduction for  $k$ -HYPVC-PARTITE is – as in [8] – the Multi-Layered PCP. While we do not explicitly construct a standalone Long Code based gadget, our reduction can be thought of as adapting the integrality gap construction of Aharoni *et al.* [1] into a Long Code based gadget in a manner that preserves the  $k$ -uniformity and  $k$ -partiteness of the integrality gap.

Such transformations of integrality gaps into Long Code based gadgets have recently been studied in the works of Raghavendra [25] and Kumar, Manokaran, Tulsiani and Vishnoi [22] which show this for a wide class of CSPs and their appropriate LP and SDP integrality gaps. These Long Code based gadgets can be combined with a Unique Games instance to yield tight UGC based hardness results, where the reduction is analyzed via the Mossel’s *Invariance Principle* [24]. Indeed, for  $k$ -HYPVC-PARTITE the work of Guruswami and Saket [13] combines the integrality gap of [1] with (a slight modification) of the approach of Kumar *et al.* [22] to obtain an optimal UGC based hardness result.

Our reduction, on the other hand, combines Long Codes with the Multi-Layered PCP instead of Unique Games and so we cannot adopt a Invariance Principle based analysis. Thus, in a flavor similar to that of [8], our analysis is via extremal combinatorics. However, our *gadget* involves several biased Long Codes with different biases and each hyperedge includes vertices from different Long Codes, unlike the construction in [8]. For our analysis, we use structural properties of a *cross-intersecting* collection of set families. A collection of set families is *cross-intersecting* if any intersection of subsets – each chosen from a different family – is large. Variants of this notion have previously been studied in extremal set theory, see for example [2]. We prove an upper bound on the measure of the smallest family in such a collection. This enables a small vertex cover (in the hypergraph of our reduction) to be *decoded* into a good labeling to the Multi-Layered PCP.

The next section defines and analyzes the above mentioned cross-intersecting set families. Section 3 defines the Multi-Layered PCP of Dinur *et al.* [8] and states their hardness for it. In Section 4 we describe our reduction and prove Theorem 1.4.

## 2 Cross-Intersecting Set Families

We use the notation  $[n] = \{1, \dots, n\}$  and  $2^{[n]} = \{F \mid F \subseteq [n]\}$ . We begin by defining cross-intersecting set families:

**Definition 2.1** A collection of  $k$  families  $\mathcal{F}_1, \dots, \mathcal{F}_k \subseteq 2^{[n]}$ , is called  $k$ -wise  $t$ -cross-intersecting if for every choice of sets  $F_i \in \mathcal{F}_i$  for  $i = 1, \dots, k$ , we have  $|F_1 \cap \dots \cap F_k| \geq t$ .

We will work with the  $p$ -biased measure on the subsets of  $[n]$ , which is defined as follows:

**Definition 2.2** Given a bias parameter  $0 < p < 1$ , we define the measure  $\mu_p$  on the subsets of  $[n]$  as:  $\mu_p(F) := p^{|F|} \cdot (1 - p)^{n - |F|}$ . The measure of a family  $\mathcal{F}$  is defined as  $\mu_p(\mathcal{F}) = \sum_{F \in \mathcal{F}} \mu_p(F)$ .

Now, we introduce an important technique for analyzing cross-intersecting families – the shift operation (see Def 4.1, pg. 1298 [12]). Given a family  $\mathcal{F}$ , define the  $(i, j)$ -shift as follows:

$$S_{ij}^{\mathcal{F}}(F) = \begin{cases} (F \cup \{i\} \setminus \{j\}) & \text{if } j \in F, i \notin F \text{ and } (F \cup \{i\} \setminus \{j\}) \notin \mathcal{F} \\ F & \text{otherwise.} \end{cases}$$

Let the  $(i, j)$ -shift of a family  $\mathcal{F}$  be  $S_{ij}(\mathcal{F}) = \{S_{ij}^{\mathcal{F}}(F) \mid F \in \mathcal{F}\}$ . Given a family  $\mathcal{F} \subseteq 2^{[n]}$ , we repeatedly apply  $(i, j)$ -shift for  $1 \leq i < j \leq n$  to  $\mathcal{F}$  until we obtain a family that is invariant under these shifts. Such a family is called a *left-shifted family* and we will denote it by  $S(\mathcal{F})$ .

The following observations about left-shifted families follow from the definition.

**Observation 2.3** Let  $\mathcal{F} \subseteq 2^{[n]}$  be a left-shifted family. Consider  $F \in \mathcal{F}$  such that  $i \notin F$  and  $j \in F$  where  $i < j$ . Then,  $(F \cup \{i\} \setminus \{j\})$  must be in  $\mathcal{F}$ .

**Observation 2.4** Given  $\mathcal{F} \subseteq 2^{[n]}$ , there is a bijection between the sets in  $\mathcal{F}$  and  $S(\mathcal{F})$  that preserves the size of the set. Thus, for any fixed  $p$ , the measures of  $\mathcal{F}$  and  $S(\mathcal{F})$  are the same under  $\mu_p$  i.e.  $\mu_p(\mathcal{F}) = \mu_p(S(\mathcal{F}))$ .

The following lemma shows that the cross-intersecting property is preserved under left-shifting.

**Lemma 2.5** Consider families  $\mathcal{F}_1, \dots, \mathcal{F}_k \subseteq 2^{[n]}$  that are  $k$ -wise  $t$ -cross-intersecting. Then, the families  $S(\mathcal{F}_1), \dots, S(\mathcal{F}_k)$  are also  $k$ -wise  $t$ -cross-intersecting.

**Proof:** Given the assumption, we will prove that  $S_{ij}(\mathcal{F}_1), \dots, S_{ij}(\mathcal{F}_k)$  are  $k$ -wise  $t$ -cross-intersecting. A simple induction would then imply the statement of the lemma.

Consider arbitrary sets  $F_i \in \mathcal{F}_i$ . By our assumption,  $|F_1 \cap \dots \cap F_k| \geq t$ . It suffices to prove that  $|S_{ij}^{\mathcal{F}_1}(F_1) \cap \dots \cap S_{ij}^{\mathcal{F}_k}(F_k)| \geq t$ . If  $j \notin F_1 \cap \dots \cap F_k$ , the claim is true since the only element being deleted is  $j$ . Thus, for all  $l \in [k]$ ,  $j \in F_l$ . If for all  $l \in [k]$ ,  $S_{ij}^{\mathcal{F}_l}(F_l) = F_l$ , the claim is trivial. Thus, let us assume wlog that  $S_{ij}^{\mathcal{F}_1}(F_1) \neq F_1$ . Thus,  $i \notin F_1$  and hence  $i \notin F_1 \cap \dots \cap F_k$ . Now, if  $i \in S_{ij}^{\mathcal{F}_1}(F_1) \cap \dots \cap S_{ij}^{\mathcal{F}_k}(F_k)$ , we get that  $j$  is replaced by  $i$  in the intersection and we are done. Thus, we can assume wlog that  $i \notin S_{ij}^{\mathcal{F}_2}(F_2)$ . This implies that  $i \notin F_2$  and  $F_2 \cup \{i\} \setminus \{j\} \in \mathcal{F}_2$ . Now consider  $F_1 \cap (F_2 \cup \{i\} \setminus \{j\}) \cap F_3 \cap \dots \cap F_k$ . Since we are picking one set from each  $\mathcal{F}_i$ , it must have at least  $t$  elements, but this intersection does not contain  $j$  and hence it is a subset of  $S_{ij}^{\mathcal{F}_1}(F_1) \cap \dots \cap S_{ij}^{\mathcal{F}_k}(F_k)$ , implying that  $|S_{ij}^{\mathcal{F}_1}(F_1) \cap \dots \cap S_{ij}^{\mathcal{F}_k}(F_k)| \geq t$ . ■

Next, we prove a key structural lemma about cross-intersecting families which states that for at least one of the families, all of its subsets have a dense prefix.

**Lemma 2.6** Let  $q_1, \dots, q_k \in (0, 1)$  be  $k$  numbers such that  $\sum_i q_i \geq 1$  and let  $\mathcal{F}_1, \dots, \mathcal{F}_k \subseteq 2^{[n]}$  be left-shifted families that are  $k$ -wise  $t$ -cross-intersecting for some  $t \geq 1$ . Then, there exists a  $j \in [k]$  such that for all sets  $F \in \mathcal{F}_j$ , there exists a positive integer  $r_F \leq n - t$  such that  $|F \cap [t + r_F]| > (1 - q_j)(t + r_F)$ .

**Proof:** Let us assume to the contrary that for every  $i \in [k]$ , there exists a set  $F_i \in \mathcal{F}_i$  such that for all  $r \geq 0$ ,  $|F_i \cap [t + r]| \leq (1 - q_i)(t + r)$ . The following combinatorial argument shows that the families  $\mathcal{F}_i$  cannot be  $k$ -wise  $t$ -cross-intersecting.

Let us construct an arrangement of balls and bins where each ball is colored with one of  $k$  colors. Create  $n$  bins labeled  $1, \dots, n$ . For each  $i$  and for every  $x \in [n] \setminus F_i$ , we place a ball with color  $i$  in the bin labeled  $x$ . Note that a bin can have several balls, but they must have distinct colors. Given such an arrangement, we can recover the sets it represents by defining  $F_i^c$  to be the set of bins that contain a ball with color  $i$ .

Our initial assumption implies that  $|F_i^c \cap [t + r]| \geq q_i(t + r)$ . Thus, there are at least  $\lceil q_i(t + r) \rceil$  balls with color  $i$  in bins labeled  $1, \dots, t + r$ . The total number of balls in bins labeled  $1, \dots, t + r$  is,

$$\sum_{i=1}^k |F_i^c \cap [t + r]| \geq \sum_{i=1}^k \lceil q_i(t + r) \rceil \geq \sum_{i=1}^k q_i(t + r) \geq (t + r) \geq r + 1,$$

where the last two inequalities follow using  $\sum_i q_i \geq 1$  and  $t \geq 1$ .

Next, we describe a procedure to manipulate the above arrangement of balls.

**for**  $r := 0$  to  $n - t$   
     **if** bin  $t + r$  is empty  
         **then if** a bin labeled from 1 to  $t - 1$  contains a ball **then** move it to bin  $t + r$   
             **else if** a bin labeled from  $t$  to  $t + r - 1$  contains two balls **then** move one of them to bin  $t + r$   
             **else** output “error”

We need the following lemma.

**Lemma 2.7** *The above procedure satisfies the following properties:*

1. *The procedure never outputs error.*
2. *At every step, any two balls in the same bin have different colors.*
3. *At step  $r$ , define  $G_i^{(r)}$  to be the set of labels of the bins that do not contain a ball of color  $i$ . Then, for all  $i \in [k]$ ,  $G_i^{(r)} \in \mathcal{F}_i$ .*
4. *After step  $r$ , the bins  $t$  to  $t + r$  have at least one ball each.*

**Proof:** 1. If it outputs error at step  $r$ , there must be at most  $r - 1$  balls in bins 1 to  $t + r$ . This is false at  $r = 0$ . Moreover, at step  $r' < r$ , we could have moved a ball only to a bin labeled in  $[t, t + r]$ . Thus, we get a contradiction.

2. Note that this is true at  $r = 0$  and a ball is only moved to an empty bin, which proves the claim.

3. Whenever we move a ball from bin  $i$  to  $j$ , we have  $i < j$ . Since  $\mathcal{F}_i$  are left-shifted, by repeated application of Observation 2.3, we get that at step  $r$ ,  $G_i^{(r)} \in \mathcal{F}_i$ .

4. Since the procedure never outputs error, at step  $r$ , if the bin  $t + r$  is empty, the procedure places a ball in it while not emptying any bin labeled between  $[t, t + r - 1]$ . This proves the claim. ■

The above lemma implies that at the end of the procedure (after  $r = n - t$ ), there is a ball in each of the bins labeled from  $[t, n]$ . Thus, the sets  $G_i = G_i^{(n-t)}$  satisfy  $\cap_i G_i \subseteq [t - 1]$  and hence  $|\cap_i G_i| \leq t - 1$ . Also, we know that  $G_i \in \mathcal{F}_i$ . Thus, the families  $\mathcal{F}_i$  cannot be  $k$ -wise  $t$ -cross-intersecting. This completes the proof of Lemma 2.6. ■

The above lemma, along with a Chernoff bound argument, shows that: Given a collection of  $k$ -wise  $t$ -cross-intersecting families, one of them must have a small measure under an appropriately chosen bias.

**Lemma 2.8** *For arbitrary  $\epsilon, \delta > 0$ , there exists some  $t = O\left(\frac{1}{\delta^2} \left(\log \frac{1}{\epsilon} + \log \left(1 + \frac{1}{2\delta^2}\right)\right)\right)$  such that the following holds: Given  $k$  numbers  $0 < q_i < 1$  such that  $\sum_i q_i \geq 1$  and  $k$  families,  $\mathcal{F}_1, \dots, \mathcal{F}_k \subseteq 2^{[n]}$ , that are  $k$ -wise  $t$ -cross-intersecting, there exists a  $j$  such that  $\mu_{1-q_j-\delta}(\mathcal{F}) < \epsilon$ .*

**Proof:** First we prove the following lemma derived from the Chernoff bound.

**Lemma 2.9** *For arbitrary  $\epsilon, \delta > 0$  and  $0 < q < 1$ , there exists some  $t = O\left(\frac{1}{\delta^2} \left(\log \frac{1}{\epsilon} + \log \left(1 + \frac{1}{2\delta^2}\right)\right)\right)$  such that the following holds:*

*Any family  $\mathcal{F} \subseteq 2^{[n]}$  that satisfies that for every  $F \in \mathcal{F}$ , there exists an integer  $r_F \geq 0$  such that  $|F \cap [t + r_F]| \geq (1 - q)(t + r_F)$  must have  $\mu_{1-q-\delta}(\mathcal{F}) < \epsilon$ .*

**Proof:** Note that  $\mu_{1-q-\delta}(\mathcal{F})$  is equal to the probability that for a random set  $F$  chosen according to  $\mu_{1-q-\delta}$  lies in  $\mathcal{F}$ . Thus,  $\mu_{1-q-\delta}(\mathcal{F})$  is bounded by the probability that for a random set  $F$  chosen according to  $\mu_{1-q-\delta}$ , there exists an  $r_F$  that satisfies  $|F \cap [t + r_F]| \geq (1 - q)(t + r_F)$ .

The Chernoff bound states that for a set of  $m$  independent bernoulli random variables  $X_i$ , with  $\Pr[X_i = 1] = 1 - q - \tau$ ,

$$\Pr \left[ \sum_{i=1}^m X_i \geq (1 - q)m \right] \leq e^{-2m\tau^2}$$

Thus, we get that for any  $r \geq 0$ ,  $\Pr[|F \cap [t + r]| \geq (1 - q)(t + r)] \leq e^{-2(t+r)\delta^2}$ . Summing over all  $r$ , we get that,

$$\mu_{1-q-\delta}(\mathcal{F}) \leq \sum_{r \geq 0} e^{-2(t+r)\delta^2} \leq \frac{e^{-2t\delta^2}}{1 - e^{-2\delta^2}} \leq e^{-2t\delta^2} \left( 1 + \frac{1}{2\delta^2} \right).$$

Thus, for  $t = \Omega \left( \frac{1}{\delta^2} \left( \log \frac{1}{\epsilon} + \log \left( 1 + \frac{1}{2\delta^2} \right) \right) \right)$ ,  $\mu_{1-q-\delta}(\mathcal{F})$  will be smaller than  $\epsilon$ . ■

We now continue with the proof of Lemma 2.8. Our  $t$  will be dictated by Lemma 2.9 and will be decided later. Consider the left-shifted families  $S(\mathcal{F}_i)$ . By Lemma 2.5, we get that these families are also  $k$ -wise  $t$ -cross-intersecting. Now, we can apply Lemma 2.6 with the given  $q_i$ 's to conclude that there must exist a  $j$  such that for all sets  $F \in S(\mathcal{F}_j)$ , there exists an  $r_F$  such that  $|F \cap [t + r_F]| > (1 - q_j)(t + r_F)$ .

Now, we can use Lemma 2.9 to conclude that if  $t$  is large enough ( $t = \Omega \left( \frac{1}{\delta^2} \left( \log \frac{1}{\epsilon} + \log \left( 1 + \frac{1}{2\delta^2} \right) \right) \right)$  suffices), then  $S(\mathcal{F}_j)$  must have measure at most  $\epsilon$  under the measure  $\mu_{1-q_j-\delta}$ , but this along with Observation 2.4 implies that  $\mu_{1-q_j-\delta}(\mathcal{F}_j) < \epsilon$ . ■

### 3 Multi-Layered PCP

In this section we describe the Multi-Layered PCP constructed in [8] and its useful properties. An instance  $\Phi$  of the Multi-Layered PCP is parametrized by integers  $L, R > 1$ . The PCP consists of  $L$  sets of variables  $X_1, \dots, X_L$ . The label set (or range) of the variables in the  $l^{\text{th}}$  set  $X_l$  is a set  $R_{X_l}$  where  $|R_{X_l}| = R^{O(L)}$ . For any two integers  $1 \leq l < l' \leq L$ , the PCP has a set of constraints  $\Phi_{l,l'}$  in which each constraint depends on one variable  $x \in X_l$  and one variable  $x' \in X_{l'}$ . The constraint (if it exists) between  $x \in X_l$  and  $x' \in X_{l'}$  ( $l < l'$ ) is denoted and characterized by a projection  $\pi_{x \rightarrow x'} : R_{X_l} \mapsto R_{X_{l'}}$ . A labeling to  $x$  and  $x'$  satisfies the constraint  $\pi_{x \rightarrow x'}$  if the projection (via  $\pi_{x \rightarrow x'}$ ) of the label assigned to  $x$  coincides with the label assigned to  $x'$ .

The following useful ‘weak-density’ property of the Multi-Layered PCP was defined in [8].

**Definition 3.1** *An instance  $\Phi$  of the Multi-Layered PCP with  $L$  layers is weakly-dense if for any  $\delta > 0$ , given  $m \geq \lceil \frac{2}{\delta} \rceil$  layers  $l_1 < l_2 < \dots < l_m$  and given any sets  $S_i \subseteq X_{l_i}$ , for  $i \in [m]$  such that  $|S_i| \geq \delta |X_{l_i}|$ ; there always exist two layers  $l_{i'}$  and  $l_{i''}$  such that the constraints between the variables in the sets  $S_{i'}$  and  $S_{i''}$  is at least  $\frac{\delta^2}{4}$  fraction of the constraints between the sets  $X_{l_{i'}}$  and  $X_{l_{i''}}$ .*

The following inapproximability of the Multi-Layered PCP was proven by Dinur et al. [8] based on the PCP Theorem ([4], [3]) and Raz’s Parallel Repetition Theorem ([26]).

**Theorem 3.2** *There exists a universal constant  $\gamma > 0$  such that for any parameters  $L > 1$  and  $R$ , there is a weakly-dense  $L$ -layered PCP  $\Phi = \cup \Phi_{l,l'}$  such that it is NP-hard to distinguish between the following two cases:*

- **YES Case:** *There exists an assignment of labels to the variables of  $\Phi$  that satisfies all the constraints.*
- **NO Case:** *For every  $1 \leq l < l' \leq L$ , not more than  $1/R^\gamma$  fraction of the constraints in  $\Phi_{l,l'}$  can be satisfied by any assignment.*



## 4 Hardness Reduction for HYPVC-PARTITE

### 4.1 Construction of the Hypergraph

Fix a  $k \geq 3$ , an arbitrarily small parameter  $\varepsilon > 0$  and let  $r = \lceil 10\varepsilon^{-2} \rceil$ . We shall construct a  $(k+1)$ -uniform  $(k+1)$ -partite hypergraph as an instance of  $(k+1)$ -HYPVC-PARTITE. Our construction will be a reduction from an instance  $\Phi$  of the Multi-Layered PCP with number of layers  $L = 32\varepsilon^{-2}$  and parameter  $R$  which shall be chosen later to be large enough. It involves creating, for each variable of the PCP, several copies of the Long Code endowed with different biased measures as explained below.

Over any domain  $T$ , a Long Code  $\mathcal{H}$  is a collection of all subsets of  $T$ , i.e.  $\mathcal{H} = 2^T$ . A bias  $p \in [0, 1]$  defines a measure  $\mu_p$  on  $\mathcal{H}$  such that  $\mu_p(v) = p^{|v|}(1-p)^{|T \setminus v|}$  for any  $v \in \mathcal{H}$ . In our construction we need several different biased measures defined as follows. For all  $j = 1, \dots, r$ , define  $q_j := \frac{2j}{rk}$ , and biases  $p_j := 1 - q_j - \varepsilon$ . Each  $p_j$  defines a biased measure  $\mu_{p_j}$  over a Long Code over any domain. Next, we define the vertices of the hypergraph.

**Vertices.** We shall denote the set of vertices by  $V$ . Consider a variable  $x$  in the layer  $X_l$  of the PCP. For  $i \in [k+1]$  and  $j \in [r]$ , let  $\mathcal{H}_{ij}^x$  be a Long Code on the domain  $R_{X_l}$  endowed with the bias  $\mu_{p_j}$ , i.e.  $\mu_{p_j}(v) = p_j^{|v|}(1-p_j)^{|R_{X_l} \setminus v|}$  for all  $v \in \mathcal{H}_{ij}^x = 2^{R_{X_l}}$ . The set of vertices corresponding to  $x$  is  $V[x] := \bigcup_{i=1}^{k+1} \bigcup_{j=1}^r \mathcal{H}_{ij}^x$ . We define the weights on vertices to be proportional to its biased measure in the corresponding Long Code. Formally, for any  $v \in \mathcal{H}_{ij}^x$ ,

$$\text{wt}(v) := \frac{\mu_{p_j}(v)}{L|X_l|r(k+1)}. \quad (1)$$

The above conveniently ensures that for any  $l \in [L]$ ,  $\sum_{x \in X_l} \text{wt}(V[x]) = 1/L$ , and  $\sum_{l \in [L]} \sum_{x \in X_l} \text{wt}(V[x]) = 1$ . In addition to the vertices for each variable of the PCP, the instance also contains  $k+1$  *dummy* vertices  $d_1, \dots, d_{k+1}$  each with a very large weight given by  $\text{wt}(d_i) := 2$  for  $i \in [k+1]$ . Clearly, this ensures that the total weight of all the vertices in the hypergraph is  $2(k+1) + 1$ . As we shall see later, the edges shall be defined in such a way that along with these weights would ensure that the maximum sized independent set shall contain all the dummy vertices. Before defining the edges we define the  $(k+1)$  partition  $(V_1, \dots, V_{k+1})$  of  $V$  to be:

$$V_i = \left( \bigcup_{l=1}^L \bigcup_{x \in X_l} \bigcup_{j=1}^r \mathcal{H}_{ij}^x \right) \cup \{d_i\}, \quad (2)$$

for all  $i = 1, \dots, k+1$ . We now define the hyperedges of the instance. In the rest of the section, the vertices shall be thought of as subsets of their respective domains.

**Hyperedges.** For every pair of variables  $x$  and  $y$  of the PCP such that there is a constraint  $\pi_{x \rightarrow y}$ , we construct edges as follows.

- (1.) Consider all permutations  $\sigma : [k+1] \mapsto [k+1]$  and sequences  $(j_1, \dots, j_k, j_{k+1})$  such that,  $j_1, \dots, j_k \in [r] \cup \{0\}$  and  $j_{k+1} \in [r]$  such that:  $\sum_{i=1}^k \mathbb{1}_{\{j_i \neq 0\}} q_{j_i} \geq 1$ .
- (2.) Add all possible hyperedges  $e$  such that for all  $i \in [k]$ :
  - (2.a) If  $j_i \neq 0$  then  $e \cap V_{\sigma(i)} =: v_{\sigma(i)} \in \mathcal{H}_{\sigma(i), j_i}^x$ , and,
  - (2.b) If  $j_i = 0$  then  $e \cap V_{\sigma(i)} = d_{\sigma(i)}$  and,
  - (2.c)  $e \cap V_{\sigma(k+1)} =: u_{\sigma(k+1)} \in \mathcal{H}_{\sigma(k+1), j_{k+1}}^y$ ,

which satisfy,

$$\pi_{x \rightarrow y} \left( \bigcap_{\substack{i: i \in [k] \\ j_i \neq 0}} v_{\sigma(i)} \right) \cap u_{\sigma(k+1)} = \emptyset. \quad (3)$$

Let us denote the hypergraph constructed above by  $G(\Phi)$ . From the construction it is clear the  $G(\Phi)$  is  $(k+1)$ -partite with partition  $V = \cup_{i \in [k+1]} V_i$ .

Note that the edges are defined in such a way that the set  $\{d_1, \dots, d_{k+1}\}$  is an independent set in the hypergraph. Moreover, since the weight of each dummy vertex  $d_i$  is 2, while total weight of all except the dummy vertices is 1, this implies that any maximum independent set  $\mathcal{I}$  contains all the dummy vertices. Thus,  $V \setminus \mathcal{I}$  is a minimum vertex cover that does not contain any dummy vertices. For convenience, the analysis of our reduction, presented in the rest of this section, shall focus on the weight of  $(\mathcal{I} \cap V) \setminus \{d_1, \dots, d_{k+1}\}$ .

## 4.2 Completeness

In the completeness case, the instance  $\Phi$  is a YES instance i.e. there is a labeling  $A$  which maps each variable  $x$  in layer  $X_l$  to an assignment in  $R_{X_l}$  for all  $l = 1, \dots, L$ , such that all the constraints of  $\Phi$  are satisfied.

Consider the set of vertices  $\mathcal{I}^*$  which satisfies the following properties:

- (1)  $d_i \in \mathcal{I}^*$  for all  $i = 1, \dots, k+1$ .
- (2) For all  $l \in [L]$ ,  $x \in X_l$ ,  $i \in [k+1]$ ,  $j \in [r]$ ,

$$\mathcal{I}^* \cap \mathcal{H}_{ij}^x = \{v \in \mathcal{H}_{ij}^x : A(x) \in v\}. \quad (4)$$

Suppose  $x$  and  $y$  are two variables in  $\Phi$  with a constraint  $\pi_{x \rightarrow y}$  between them. Consider any  $v \in \mathcal{I}^* \cap V[x]$  and  $u \in \mathcal{I}^* \cap V[y]$ . The above construction of  $\mathcal{I}^*$  along with the fact that the labeling  $A$  satisfies the constraint  $\pi_{x \rightarrow y}$  implies that  $A(x) \in v$  and  $A(y) \in u$  and  $A(y) \in \pi_{x \rightarrow y}(v) \cap u$ . Therefore, Equation (3) of the construction is not satisfied by the vertices in  $\mathcal{I}^*$ , and so  $\mathcal{I}^*$  is an independent set in the hypergraph. By Equation (4), the fraction of the weight of the Long Code  $\mathcal{H}_{ij}^x$  which lies in  $\mathcal{I}^*$  is  $p_j$ , for any variable  $x$ ,  $i \in [k+1]$  and  $j \in [r]$ . Therefore,

$$\frac{\text{wt}(\mathcal{I}^* \cap V[x])}{\text{wt}(V[x])} = \frac{1}{r} \sum_{j=1}^r p_j = 1 - \frac{1}{k} \left( 1 + \frac{1}{r} \right) - \varepsilon, \quad (5)$$

by our setting of  $p_j$  in Section 4.1. The above yields that

$$\text{wt}(\mathcal{I}^* \cap (V \setminus \{d_1, \dots, d_{k+1}\})) = 1 - \frac{1}{k} \left( 1 + \frac{1}{r} \right) - \varepsilon \geq 1 - \frac{1}{k} - 2\varepsilon, \quad (6)$$

for a small enough value of  $\varepsilon > 0$  and our setting of the parameter  $r$ .

## 4.3 Soundness

For the soundness analysis we have that  $\Phi$  is a NO instance as given in Theorem 3.2 and we wish to prove that the size of the maximum independent set in  $G(\Phi)$  is appropriately small. For a contradiction, we assume that there is a maximum independent set  $\mathcal{I}$  in  $G(\Phi)$  such that,

$$\text{wt}(\mathcal{I} \cap (V \setminus \{d_1, \dots, d_{k+1}\})) \geq 1 - \frac{k}{2(k+1)} + \varepsilon. \quad (7)$$



Define the set of variables  $X'$  to be as follows:

$$X' := \left\{ x \text{ a variable in } \Phi : \frac{\text{wt}(\mathcal{I} \cap V[x])}{\text{wt}(V[x])} \geq 1 - \frac{k}{2(k+1)} + \frac{\varepsilon}{2} \right\}. \quad (8)$$

An averaging argument shows that  $\text{wt}(\cup_{x \in X'} V[x]) \geq \varepsilon/2$ . A further averaging implies that there are  $\frac{\varepsilon}{4}L = \frac{8}{\varepsilon}$  layers of  $\Phi$  such that  $\frac{\varepsilon}{4}$  fraction of the variables in each of these layers belong to  $X'$ . Applying the Weak Density property of  $\Phi$  given by Definition 3.1 and Theorem 3.2 yields two layers  $X_{l'}$  and  $X_{l''}$  ( $l' < l''$ ) such that  $\frac{\varepsilon^2}{64}$  fraction of the constraints between them are between variables in  $X'$ . The rest of the analysis shall focus on these two layers and for convenience we shall denote  $X' \cap X_{l'}$  by  $X$  and  $X' \cap X_{l''}$  by  $Y$ , and denote the respective label sets by  $R_X$  and  $R_Y$ .

Consider any variable  $x \in X$ . For any  $i \in [k+1], j \in [r]$ , call a Long Code  $\mathcal{H}_{ij}^x$  *significant* if  $\mu_{p_j}(\mathcal{I} \cap \mathcal{H}_{ij}^x) \geq \frac{\varepsilon}{2}$ . From Equation (8) and an averaging argument we obtain that,

$$|\{(i, j) \in [k+1] \times [r] : \mathcal{H}_{ij}^x \text{ is significant}\}| \geq \left(1 - \frac{k}{2(k+1)}\right)(r(k+1)) = \frac{rk}{2} + r. \quad (9)$$

Using an analogous argument we obtain a similar statement for every variable  $y \in Y$  and corresponding Long Codes  $\mathcal{H}_{ij}^y$ . The following structural lemma follows from the above bound.

**Lemma 4.1** *Consider any variable  $x \in X$ . Then there exists a sequence  $(j_1, \dots, j_{k+1})$  with  $j_i \in [r] \cup \{0\}$  for  $i \in [k+1]$ ; such that the Long Codes  $\{\mathcal{H}_{i,j_i}^x \mid i \in [k+1] \text{ where } j_i \neq 0\}$ , are all significant. Moreover,*

$$\sum_{i=1}^{k+1} j_i \geq \frac{rk}{2} + r. \quad (10)$$

**Proof:** For all  $i \in [k+1]$  choose  $j_i$  as follows: if none of the Long Codes  $\mathcal{H}_{ij}^x$  for  $j \in [r]$  are *significant* then let  $j_i := 0$ , otherwise let  $j_i := \max\{j \in [r] : \mathcal{H}_{ij}^x \text{ is significant}\}$ . It is easy to see that  $j_i$  is an upper bound on the number of significant Long Codes of the form  $\mathcal{H}_{ij}^x$ . Therefore,

$$\sum_{i=1}^{k+1} j_i \geq |\{(i, j) \in [k+1] \times [r] : \mathcal{H}_{ij}^x \text{ is significant}\}| \geq \frac{rk}{2} + r \quad (\text{From Equation (9)}) \quad (11)$$

which proves the lemma. ■

Next we define the decoding procedure to define a label for any given variable  $x \in X$ .

#### 4.3.1 Labeling for variable $x \in X$

The label  $A(x)$  for each variable  $x \in X$  is chosen independently via the following three step (randomized) procedure.

Step 1. Choose a sequence  $(j_1, \dots, j_{k+1})$  yielded by Lemma 4.1 applied to  $x$ .

Step 2. Choose an element  $i_0$  uniformly at random from  $[k+1]$ .

Before describing the third step of the procedure we require the following lemma.

**Lemma 4.2** *There exist vertices  $v_i \in \mathcal{I} \cap \mathcal{H}_{i,j_i}^x$  for every  $i : i \in [k+1] \setminus \{i_0\}, j_i \neq 0$ , and an integer  $t := t(\varepsilon)$  satisfying:*

$$\left| \bigcap_{\substack{i: i \in [k+1] \setminus \{i_0\}, \\ j_i \neq 0}} v_i \right| < t. \quad (12)$$

**Proof:** Since  $j_{i_0} \leq r$  it is easy to see,

$$\sum_{i \in [k+1] \setminus \{i_0\}} j_i \geq \frac{rk}{2} \Rightarrow \sum_{\substack{i \in [k+1] \setminus \{i_0\}, \\ j_i \neq 0}} q_{j_i} \geq 1. \quad (13)$$

Moreover, since the sequence  $(j_1, \dots, j_{k+1})$  was obtained by Lemma 4.1 applied to  $x$ , we know that  $\mu_{p_{j_i}}(\mathcal{I} \cap \mathcal{H}_{i j_i}^x) \geq \frac{\varepsilon}{2}$ ,  $\forall i : i \in [k+1] \setminus \{i_0\}, j_i \neq 0$ . Combining this with Equation (13) and Lemma 2.8 we obtain that for some integer  $t := t(\varepsilon)$  the collection of set families  $\{\mathcal{H}_{i j_i}^x : i \in [k+1] \setminus \{i_0\}, j_i \neq 0\}$  is not  $k'$ -wise  $t$ -cross-intersecting, where  $k' = |\{i \in [k+1] \setminus \{i_0\} : j_i \neq 0\}|$ . This proves the lemma. ■

The third step of the labeling procedure is as follows:

Step 3. Apply Lemma 4.2 to obtain the vertices  $v_i \in \mathcal{I} \cap \mathcal{H}_{i j_i}^x$  for every  $i : i \in [k+1] \setminus \{i_0\}, j_i \neq 0$  satisfying Equation (12). Define  $B(x)$  as,

$$B(x) := \bigcap_{\substack{i \in [k+1] \setminus \{i_0\}, \\ j_i \neq 0}} v_i, \quad (14)$$

noting that  $|B(x)| < t$ . Assign a random label from  $B(x)$  to the variable  $x$  and call the assigned label  $A(x)$ .

#### 4.3.2 Labeling for variable $y \in Y$

After labeling the variables  $x \in X$  via the procedure above, we construct a labeling  $A(y)$  for any variable  $y \in Y$  by defining,

$$A(y) := \operatorname{argmax}_{a \in R_Y} |\{x \in X \cap N(y) \mid a \in \pi_{x \rightarrow y}(B(x))\}|, \quad (15)$$

where  $N(y)$  is the set of all variables that have a constraint with  $y$ . The above process selects a label for  $y$  which lies in maximum number of projections of  $B(x)$  for variables  $x \in X$  which have a constraint with  $y$ . The rest of this section is devoted to lower bounding the number of constraints satisfied by the labeling process, and thus obtain a contradiction to the fact that  $\Phi$  is a NO instance.

#### 4.3.3 Lower bounding the number of satisfied constraints

Fix a variable  $y \in Y$ . Let  $U(y) := X \cap N(y)$ , i.e. the variables in  $X$  which have a constraint with  $y$ . Further, define the set  $P(y) \subseteq [k+1]$  as follows,

$$P(y) = \{i \in [k+1] \mid \exists j \in [r] \text{ such that } \mu_{p_j}(\mathcal{I} \cap \mathcal{H}_{i j}^y) \geq \varepsilon/2\}. \quad (16)$$

In other words,  $P(y)$  is the set of all those indices in  $[k+1]$  such that there is a *significant* Long Code corresponding to each of them. Applying Equation (9) to  $y$  we obtain that there at least  $\frac{r(k+2)}{2}$  *significant* Long Codes corresponding to  $y$ , and therefore  $|P(y)| \geq \frac{k+1}{2} \geq 1$ . Next we define subsets of  $U(y)$  depending on the outcome of Step 2 in the labeling procedure for variables  $x \in U(y)$ . For  $i \in [k+1]$  define,

$$U(i, y) := \{x \in U(y) \mid i \text{ was chosen in Step 2 of the labeling procedure for } x\}, \quad (17)$$

and,

$$U^*(y) := \bigcup_{i \in P(y)} U(i, y). \quad (18)$$

Note that  $\{U(i, y)\}_{i \in [k+1]}$  is a partition of  $U(y)$ . Also, since  $|P(y)| \geq \frac{k+1}{2}$  and the labeling procedure for each variable  $x$  chooses the index in Step 2 uniformly and independently at random we have,

$$\mathbb{E}[|U^*(y)|] \geq \frac{|U(y)|}{2}, \quad (19)$$

where the expectation is over the random choice of the indices in Step 2 of the labeling procedure for all  $x \in U(y)$ . Before continuing we need the following simple lemma (proved as Claim 5.4 in [8]).

**Lemma 4.3** *Let  $A_1, \dots, A_N$  be a collection of  $N$  sets, each of size at most  $T \geq 1$ . If there are not more than  $D$  pairwise disjoint sets in the collection, then there is an element that is contained in at least  $\frac{N}{TD}$  sets.*

Now consider any  $i' \in P(y)$  such that  $U(i', y) \neq \emptyset$  and a variable  $x \in U(i', y)$ . Since  $i' \in P(y)$  there is a significant Long Code  $\mathcal{H}_{i', j'}^y$  for some  $j' \in [r]$ . Furthermore, since  $\mathcal{I}$  is an independent set there cannot be a  $u \in \mathcal{I} \cap \mathcal{H}_{i', j'}^y$  such that  $\pi_{x \rightarrow y}(B(x)) \cap u = \emptyset$ , otherwise the following set of  $k+1$  vertices,

$$\{v_i \mid i \in [k+1] \setminus \{i'\}, j_i \neq 0\} \cup \{d_i \mid i \in [k+1] \setminus \{i'\}, j_i = 0\} \cup \{u\}$$

form an edge in  $\mathcal{I}$ , where  $v_i, j_i$  ( $i \in [k+1]$ ) are as constructed in the labeling procedure for  $x$ .

Consider the collection of sets  $\pi_{x \rightarrow y}(B(x))$  for all  $x \in U(i', y)$ . Clearly each set is of size less than  $t$ . Let  $D$  be the maximum number of disjoint sets in this collection. Each disjoint set independently reduces the measure of  $\mathcal{I} \cap \mathcal{H}_{i', j'}^y$  by a factor of  $(1 - (1 - p_{j'})^t)$ . However, since  $\mu_{p_{j'}}(\mathcal{I} \cap \mathcal{H}_{i', j'}^y)$  is at least  $\frac{\varepsilon}{2}$ , this implies that  $D$  is at most  $\log(\frac{\varepsilon}{2}) / \log(1 - (2/rk)^t)$ , since  $p_{j'} \leq 1 - \frac{2}{rk}$ . Moreover, since  $t$  and  $r$  depends only on  $\varepsilon$ , the upper bound on  $D$  also depends only on  $\varepsilon$ .

Therefore by Lemma 4.3, there is an element  $a \in R_Y$  such that  $a \in \pi_{x \rightarrow y}(B(x))$  for at least  $\frac{1}{Dt}$  fraction of  $x \in U(i', y)$ . Noting that this bound is independent of  $j'$  and that  $\{U(i', y)\}_{i' \in P(y)}$  is a partition of  $U^*(y)$ , we obtain that there is an element  $a \in R_Y$  such that  $a \in \pi_{x \rightarrow y}(B(x))$  for  $\frac{1}{(k+1)Dt}$  fraction of  $x \in U^*(y)$ . Therefore, in Step 3 of the labeling procedure when a label  $A(x)$  is chosen uniformly at random from  $B(x)$ , in exception,  $a = \pi_{x \rightarrow y}(A(x))$  for  $\frac{1}{(k+1)Dt^2}$  fraction of  $x \in U^*(y)$ . Combining this with Equation (19) gives us that there is a labeling to the variables in  $X$  and  $Y$  which satisfies  $\frac{1}{2(k+1)Dt^2}$  fraction of the constraints between variables in  $X$  and  $Y$  which is in turn at least  $\frac{\varepsilon^2}{64}$  fraction of the constraints between the layers  $X_{\nu'}$  and  $X_{\nu''}$ . Since  $D$  and  $t$  depend only on  $\varepsilon$ , choosing the parameter  $R$  of  $\Phi$  to be large enough we obtain a contradiction to our supposition on the lower bound on the size of the independent set. Therefore in the Soundness case, any for any independent set  $\mathcal{I}$ ,

$$\text{wt}(\mathcal{I} \cap (V \setminus \{d_1, \dots, d_{k+1}\})) \leq 1 - \frac{k}{2(k+1)} + \varepsilon.$$

Combining the above with Equation (6) of the analysis in the Completeness case yields a factor  $\frac{k^2}{2(k+1)} - \delta$  (for any  $\delta > 0$ ) hardness for approximating  $(k+1)$ -HYPVC-PARTITE.

Thus, we obtain a factor  $\frac{k}{2} - 1 + \frac{1}{2k} - \delta$  hardness for approximating  $k$ -HYPVC-PARTITE.

## References

- [1] R. Aharoni, R. Holzman, and M. Krivelevich. On a theorem of Lovász on covers in  $r$ -partite hypergraphs. *Combinatorica*, 16(2):149–174, 1996.
- [2] Noga Alon and Eyal Lubetzky. Uniformly cross intersecting families. *Combinatorica*, 29(4):389–431, 2009.

- [3] S. Arora, C. Lund, R. Motwani, M. Sudan, and M. Szegedy. Proof verification and the hardness of approximation problems. *J. ACM*, 45(3):501–555, 1998.
- [4] S. Arora and S. Safra. Probabilistic checking of proofs: A new characterization of NP. *J. ACM*, 45(1):70–122, 1998.
- [5] Per Austrin, Subhash Khot, and Muli Safra. Inapproximability of vertex cover and independent set in bounded degree graphs. In *IEEE Conference on Computational Complexity*, pages 74–80, 2009.
- [6] Nikhil Bansal and Subhash Khot. Inapproximability of hypergraph vertex cover and applications to scheduling problems. In *ICALP*, pages 250–261, 2010.
- [7] I. Dinur, V. Guruswami, and S. Khot. Vertex cover on  $k$ -uniform hypergraphs is hard to approximate within factor  $(k-3-\epsilon)$ . ECCC Technical Report TR02-027, 2002.
- [8] I. Dinur, V. Guruswami, S. Khot, and O. Regev. A new multilayered PCP and the hardness of hypergraph vertex cover. In *Proc. 35<sup>th</sup> ACM STOC*, pages 595–601, 2003.
- [9] I. Dinur and S. Safra. The importance of being biased. In *Proc. 34<sup>th</sup> ACM STOC*, pages 33–42, 2002.
- [10] T. Feder, R. Motwani, L. O’Callaghan, R. Panigrahy, and D. Thomas. Online distributed predicate evaluation. Technical Report, Stanford University, 2003.
- [11] G. Gottlob and P. Senellart. Schema mapping discovery from data instances. *J. ACM*, 57(2), January 2010.
- [12] R. L. Graham, M. Grötschel, and L. Lovász, editors. *Handbook of combinatorics (vol. 2)*. MIT Press, Cambridge, MA, USA, 1995.
- [13] Venkatesan Guruswami and Rishi Saket. On the inapproximability of vertex cover on  $k$ -partite  $k$ -uniform hypergraphs. In *ICALP*, pages 360–371, 2010.
- [14] J. Håstad. Some optimal inapproximability results. *J. ACM*, 48(4):798–859, 2001.
- [15] E. Halperin. Improved approximation algorithms for the vertex cover problem in graphs and hypergraphs. *SIAM J. Comput.*, 31(5):1608–1623, 2002.
- [16] J. Holmerin. Improved inapproximability results for vertex cover on  $k$ -uniform hypergraphs. In *Proc. ICALP*, pages 1005–1016, 2002.
- [17] L. Ilie, R. Solis-Oba, and S. Yu. Reducing the size of NFAs by using equivalences and preorders. In *Proc. CPM*, pages 310–321, 2005.
- [18] George Karakostas. A better approximation ratio for the vertex cover problem. In *ICALP*, pages 1043–1050, 2005.
- [19] Richard M. Karp. Reducibility among combinatorial problems. *Complexity of Computer Computations*, pages 85–103, 1972.
- [20] S. Khot. On the power of unique 2-prover 1-round games. In *Proc. 34<sup>th</sup> ACM STOC*, pages 767–775, 2002.
- [21] S. Khot and O. Regev. Vertex cover might be hard to approximate to within 2-epsilon. *J. Comput. Syst. Sci.*, 74(3):335–349, 2008.

- [22] A. Kumar, R. Manokaran, M. Tulsiani, and N. K. Vishnoi. On the optimality of a class of LP-based algorithms. Manuscript, 2009.
- [23] L. Lovász. On minimax theorems of combinatorics. *Doctoral Thesis, Mathematiki Lapok*, 26:209–264, 1975.
- [24] E. Mossel. Gaussian bounds for noise correlation of functions and tight analysis of long codes. In *Proc. 49<sup>th</sup> IEEE FOCS*, pages 156–165, 2008.
- [25] P. Raghavendra. Optimal algorithms and inapproximability results for every CSP? In *Proc. 40<sup>th</sup> ACM STOC*, pages 245–254, 2008.
- [26] R. Raz. A parallel repetition theorem. *SIAM J. Comput.*, 27(3):763–803, 1998.
- [27] L. Trevisan. Non-approximability results for optimization problems on bounded degree instances. In *Proc. 33<sup>rd</sup> ACM STOC*, pages 453–461, 2001.

## A LP Integrality Gap for $k$ -HYPVC-PARTITE

This section describes the  $\frac{k}{2} - o(1)$  integrality gap construction of Aharoni *et al.* [1] for the standard LP relaxation for  $k$ -HYPVC-PARTITE. The hypergraph that is constructed is unweighted.

Let  $r$  be a (large) positive integer. The vertex set  $V$  of the hypergraph is partitioned into subsets  $V_1, \dots, V_k$  where, for all  $i = 1, \dots, k$ ,

$$V_i = \{x_{ij} \mid j = 1, \dots, r\} \cup \{y_{il} \mid l = 1, \dots, rk + 1\}. \quad (20)$$

Before we define the hyperedges, for convenience we shall define the LP solution. The LP values of the vertices are as given by the function  $h : V \mapsto [0, 1]$  as follows: for all  $i = 1, \dots, k$ ,

$$\begin{aligned} h(x_{ij}) &= \frac{2j}{rk}, & \forall j = 1, \dots, r \\ h(y_{il}) &= 0, & \forall l = 1, \dots, rk + 1. \end{aligned}$$

The set of hyperedges is naturally defined to be the set of all possible hyperedges, choosing exactly one vertex from each  $V_i$  such that the sum of the LP values of the corresponding vertices is at least 1. Formally,

$$E = \{e \subseteq V \mid \forall i \in [k], |e \cap V_i| = 1 \text{ and } \sum_{v \in e} h(v) \geq 1\}. \quad (21)$$

Clearly the graph is  $k$ -uniform and  $k$ -partite with  $\{V_i\}_{i \in [k]}$  being the  $k$ -partition of  $V$ .

The value of the LP solution is

$$\sum_{v \in V} h(v) = k \sum_{j \in [r]} \frac{2j}{rk} = r + 1. \quad (22)$$

Now let  $V'$  be a minimum vertex cover in the hypergraph. To lower bound the size of the minimum vertex cover, we first note that the set  $\{v \in V \mid h(v) > 0\}$  is a vertex cover of size  $rk$ , and therefore  $|V'| \leq rk$ . Also, for any  $i \in [k]$  the vertices  $\{y_{il}\}_{l \in [rk+1]}$  have the same neighborhood. Therefore, we can assume that  $V'$  has no vertex  $y_{il}$ , otherwise it will contain at least  $rk + 1$  such vertices.

For all  $i \in [k]$  let define indices  $j_i \in [r] \cup \{0\}$  as follows:

$$j_i = \begin{cases} 0 & \text{if: } \forall j \in [r], x_{ij} \in V', \\ \max \{j \in [r] \mid x_{ij} \notin V'\} & \text{otherwise.} \end{cases} \quad (23)$$

It is easy to see that since  $V'$  is a vertex cover,

$$\sum_{i \in [k]} h(x_{ij_i}) < 1,$$

which implies,

$$\sum_{i \in [k]} j_i < \frac{rk}{2}.$$

Also, the size of  $V'$  is lower bounded by  $\sum_{i \in [k]} (r - j_i)$ . Therefore,

$$|V'| \geq \sum_{i \in [k]} (r - j_i) \geq rk - \sum_{i \in [k]} j_i \geq rk - \frac{rk}{2} = \frac{rk}{2}. \quad (24)$$

The above combined with the value of the LP solution yields an integrality gap of  $\frac{rk}{2(r+1)} \geq \frac{k}{2} - o(1)$  for large enough  $r$ .